

# SCIENTIFIC REPORTS

OPEN

## Spatial dynamics of TB within a highly urbanised Asian metropolis using point patterns

Sourav Das<sup>1</sup>, Alex R. Cook<sup>2</sup>, Win Wah<sup>2</sup>, Khin Mar Kyi Win<sup>3</sup>, Cynthia Bin Eng Chee<sup>3</sup>, Yee Tang Wang<sup>3</sup> & Li Yang Hsu<sup>2,3</sup> 

Received: 27 June 2016

Accepted: 31 January 2017

Published online: 24 February 2017

Singapore is a high-income country in a region with a high prevalence of tuberculosis. The Singapore Tuberculosis (TB) Elimination Program (STEP) was set up in 1997, and the better surveillance and clinical management practices initiated under STEP led to a decade-long decline in the incidence levels. However, incidence rates started to rise again since 2008. The reasons for this rise are unclear. This study involved a spatial analysis of the epidemiology of TB among Singapore residents. More than 30 000 cases reported during 1995–2011 and their residential addresses were analysed for spatial risk and spatial clustering, using spatial point pattern methodology. The principal factor responsible for the increasing resident TB incidence in Singapore is the changing age profile of the population. In particular the burgeoning population aged above 65 years accounts for the increase in reported cases. Singapore's population has one of the world's lowest fertility and mortality rates, and the elderly population is projected to grow substantially over the next few decades. Tuberculosis rates may therefore continue to rise even with static or improving case management and surveillance.

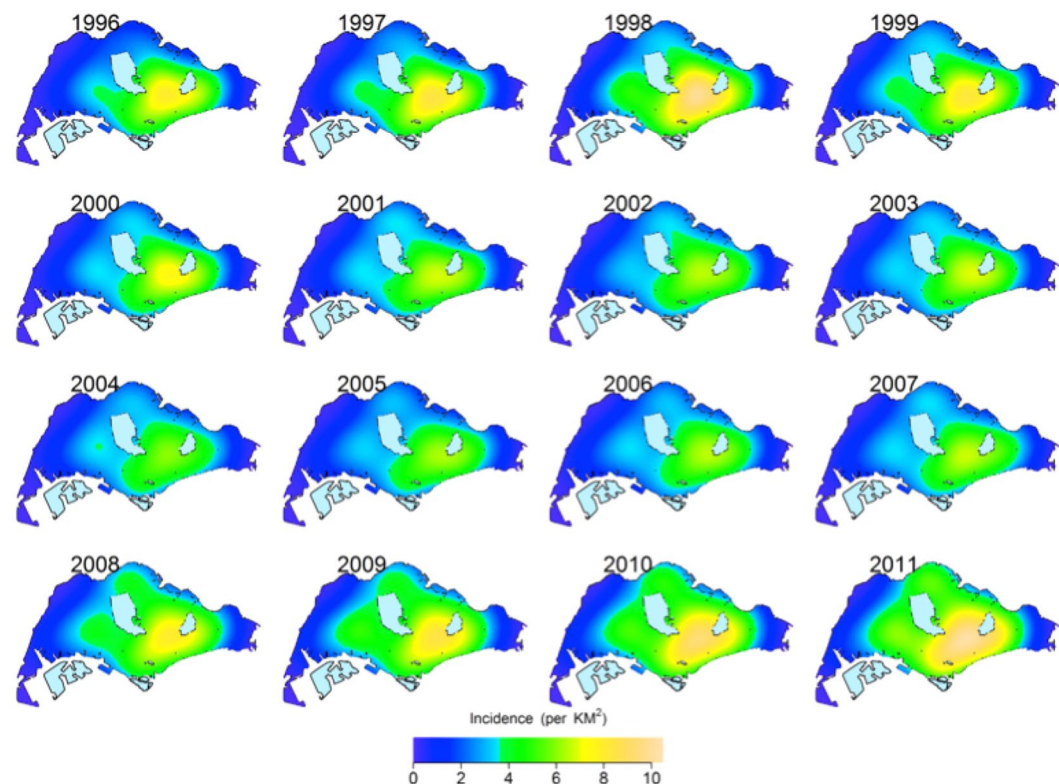
Tuberculosis (TB) has been a leading cause of death in human societies for decades. In 1993 the World Health Organization declared TB to be a global health emergency<sup>1,2</sup>. Its reduction was later included as one of the challenges in the global Millennium developmental agenda<sup>3,4</sup> to be achieved by 2015<sup>5</sup>. In 1995 an annual surveillance program was launched under the auspices of the World Health Organization (WHO). During the past two decades there have been substantial accomplishments. In particular, the mortality rate has reduced by 47%<sup>5</sup>, and TB incidence is 18% lower than the levels of 2000. However, challenges remain. In its latest report<sup>5</sup>, the WHO notes that in 2014 an estimated 9.6 million fell ill with TB. Of these, 6 million cases were new patients. Together with HIV, TB continues to be a major cause of death (1.2 million reported deaths in 2014) globally.

Of the new cases, 58% are estimated to be in the South East Asia or Western Pacific regions. In this study we investigate the epidemiology of TB among the susceptible resident population of south-east Asia's most economically advanced nation, Singapore, situated at the end of the Malay Peninsula. The city state's location in one of the high TB burden zones of the world makes her vulnerable to tuberculosis. As the global urban population crossed 50% for the first time in 2014, largely driven by developing nations<sup>6</sup>, Singapore's TB epidemiology may have important lessons for densely populated urban areas of high TB burden countries such as India (23% of global cases), China (10%) and Indonesia (10%).

Incidence rates among Singapore residents declined from 300 (per 100,000pa) in the early 1960s to 106 (per 100,000pa) in the mid-1980s<sup>7–9</sup>. In 1997, the Singapore Tuberculosis Elimination Programme (STEP)<sup>7</sup> was launched to address a decade-long stagnation in TB rates at 50–55 (per 100,000pa) during 1987 to 1997. Thereafter, TB incidence rates among Singapore residents declined from 58 (per 100,000pa) in 1998 to a historic low of 35 (per 100,000pa) in 2006 and 2007. However, since 2008 the TB incidence has begun to rise to around 39 per 100,000 population per year. The reasons for this surge, which contrasts with global improvements, remain unclear since there has been no association study on the influence of the demographic characteristics on TB in Singapore.

In the present study, we analyse more than 30,000 cases reported between 1995 and 2011. The aim of this paper is twofold: first, to analyse spatial risk of TB prevalence in Singapore and investigate geographically structured

<sup>1</sup>School of Mathematics, University of Bristol, Bristol, UK. <sup>2</sup>Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. <sup>3</sup>Tuberculosis Control Unit, Singapore TB Elimination Program, Ministry of Health, Singapore, Singapore. Correspondence and requests for materials should be addressed to L.Y.H. (email: [hsuliyang@gmail.com](mailto:hsuliyang@gmail.com))



**Figure 1.** Interpolated annual spatial intensity maps of TB risk in Singapore based on all reported cases. Purple implies lowest intensity and dark yellow signify high intensity areas. Plots were created in statistical program R<sup>22</sup> using the package *spatstat*<sup>19</sup>. Administrative and national political coordinates of Singapore were provided, for academic research and publication, by the Singapore Land Authority (SLA)<sup>24</sup>.

factors among residents at planning area levels of the city state, developing spatial risk maps for public health interventions. Second, we assess demographic factors associated with the rise in reported TB incidence among Singapore residents. A detailed epidemiology of non-resident cases will be reported separately.

To this end, we apply statistical point pattern methodology, to the residential addresses of patients, to investigate the following: 1a. Does the spatial spread of TB indicate heterogeneity? 1b. Has the spatial trend changed significantly over time? If so which are the zones with higher prevalence? 2. What is the major demographic factor behind the increase of TB among resident Singaporeans? 3. Are there indications of transmission between residents and non-residents? Non-resident cases (and their residential addresses used here) exclude individuals who have lived in Singapore for less than three months. This censors short term visitors and medical tourists.

## Results

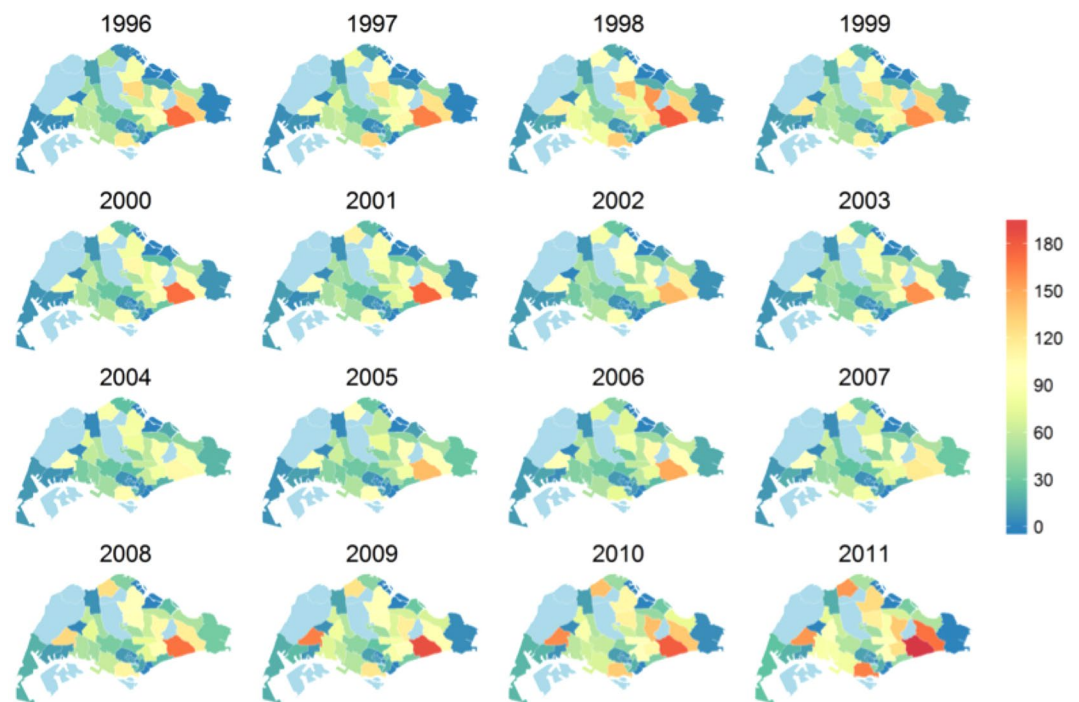
The estimated spatial intensity of TB cases across Singapore is presented in Fig. 1, indicating substantial spatial heterogeneity across the island and the evolution of spatial incidence with time. The intensity of TB cases increased in the east and southern parts of the city, between 1995 and 1999, followed by a period of decline during 2000–2007 and showed a resurgence between 2008 and 2011.

Figure 2 shows the spatial intensity map of TB at the planning area level. A few specific planning areas have noteworthy patterns: Bedok in the south east has consistently featured as the area with highest TB risk, followed by Tampines and Hougang, also in the east. Woodlands in the north and Bukit Merah in the south also have relatively high disease intensity, while Jurong West has seen a sharp increase in TB incidence.

The spatial clustering of TB, as measured via the inhomogeneous  $K$ -function, is presented in Fig. 3, stratified by year. The  $K$  function was separately plotted for *all* cases as well as for two demographic constituents, residents and non-residents. Figure 3 also presents inter-group (resident cases in the neighbourhood of non-resident cases) clustering using the cross- $K_c(\cdot)$  function.

Within each demographic grouping it can be seen that TB case residences are more clustered than they would be due to chance alone, assuming a heterogeneous intensity function,  $\lambda(s)$ , for all distances considered up to 2 km. This may reflect heterogeneity in underlying risk due to differences in age structure or the environment, for example, or be due to stochastic clustering; as noted by Bartlett<sup>10</sup> these cannot be distinguished using a single point pattern.

When all cases are considered together, in the top left panel, we note that the extent of clustering decreased from the 1990s to the later 2000s, indicating that while the incidence was generally higher in 1998 than in 2009 (Fig. 1), cases became less clustered over time. The cross- $K_c(\cdot)$  function plot does not indicate significant residential clustering of resident cases around non-resident cases or vice-versa, suggesting that disease among the



**Figure 2.** Estimated case count of TB cases in 44 planning areas of Singapore. Counts were computed by integrating the estimated spatial intensity over the planning area polygons. The plots capture higher risk of cases in Bedok and the surge of cases in Jurong West. Plots were created in statistical program R<sup>22</sup> using the package *spatstat*<sup>19</sup>. Administrative and national political coordinates of Singapore were provided, for academic research and publication, by the Singapore Land Authority (SLA)<sup>24</sup>.

two groups is close to being independent. Figure 3 also presents the cross K function between residents and non-residents in 2000 and 2010 for three zones with high (Bedok, Jurong West, Tampines) and low (Queenstown, Rochor, Sengkang) TB incidence. There is no evidence of clustering between the two demographic groups, with the cross K function indicating that cluster levels are close to what would be expected by two processes that are either independent or have a slight negative (i.e. repulsive) relationship.

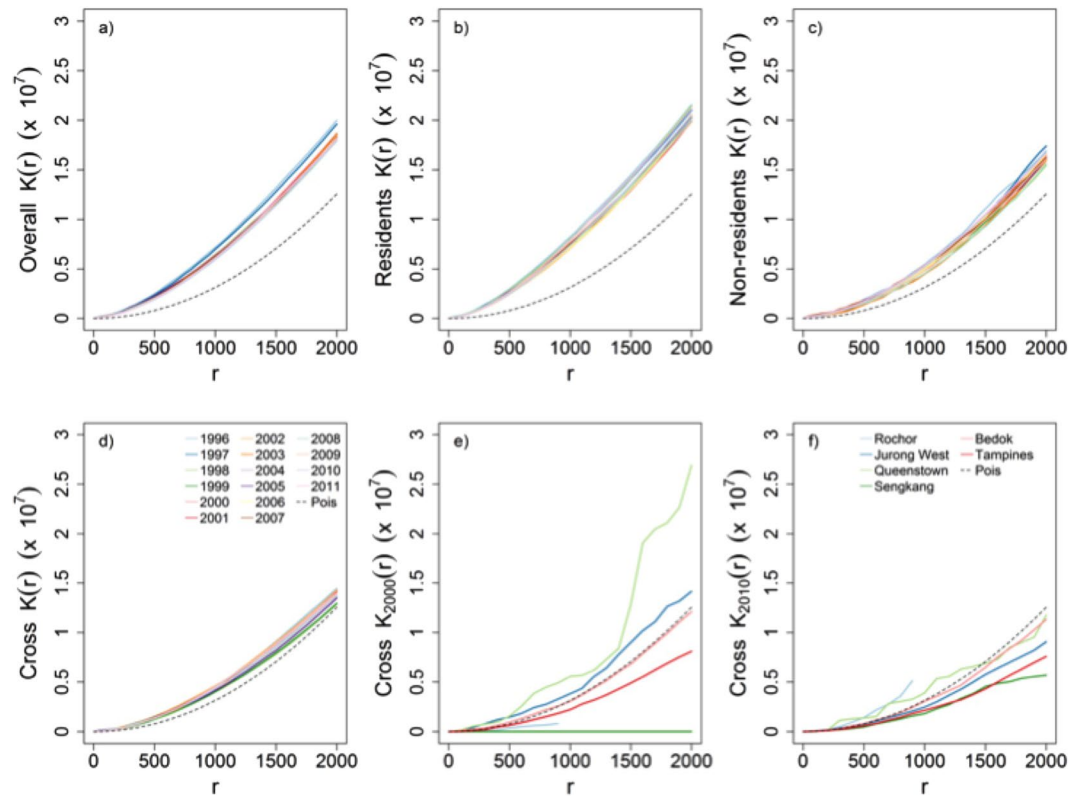
Figure 4 shows the  $p$ -values for all distinct pairs of years using the spatial Hotelling's  $T^2$  test, to compare resident TB incidence rates (per 10,000) at planning area levels. From 1996 to 2011 the  $p$ -value matrix displays a roughly uniform distribution, with only a slight skew towards significant differences: 15% of  $p$ -values are significant at the 5% and 5% at the 1% level, implying there is little evidence of a change in the population adjusted annual TB incidence among residents. Note that in 1995 – the first year of residential address based notification – many cases remained unreported, and this year has most of the significant differences present in this analysis. This has led to statistically significant difference of annual incidence with all other years, across planning areas. This is shown in Supplementary Figure 1.

Figures 5 and 6 show the reported incidence rates of TB among residents, in Singapore, over six different age groups in 2000 and in 2010, respectively. The figures indicate that, for Singapore residents, incidence is highest in the age group of over 65 year olds, and that incidence within this age group fell from 20–30 reported TB cases (per 10,000 population) in 2000 to below 15 cases (per 10,000) in most planning areas, including all major planning areas such as Bedok, Bukit Merah and Jurong West.

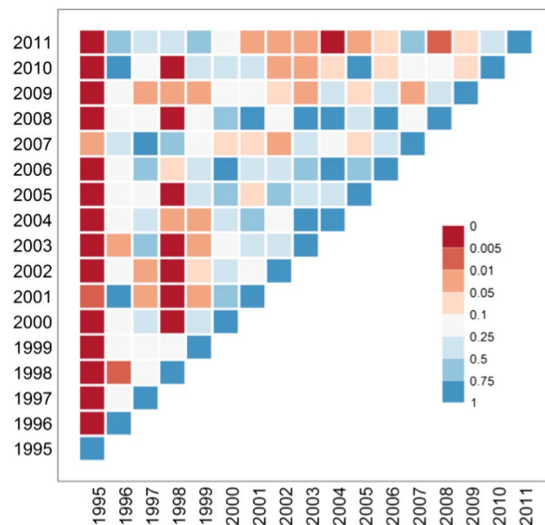
In Fig. 7 we show the annual time series of resident TB cases from 2000 to 2010 for the six age groups, adjusted for population, across ten major planning areas. Together with Figs 5 and 6 the panels of Fig. 7 corroborate two principal factors: firstly there is a clear gradient of TB incidence from lowest levels (close to 0 per 10,000) among the youngest age group 0–24 to highest levels (of above 16 cases per 10,000) the senior age group of above 65. Secondly, on adjusting for susceptible population the annual trends of TB incidence for almost all planning areas remain stable across all age groups.

## Discussion

TB incidence in Singapore is lower than in neighbouring countries, but higher than other countries with similar levels of economic development, and incidence rates have been rising despite earlier successes<sup>9</sup>. In an earlier paper, we had investigated the temporal evolution of TB incidence in Singapore<sup>11</sup>. In this study, using more than 22,000 cases reported during 1995–2011, we have analysed the spatial variation of TB incidence across the city state, among residents. The period 1995–2002 saw cases concentrated in the south eastern planning areas of Singapore. Bedok consistently had the highest concentration of around 10 cases per square km with high to medium case density (between 4–8 cases per square km) in the surrounding areas of Tampines, Hougang and Ang Mo Kio. The western planning areas had low spatial risk, reflecting the relatively lower residential density in these

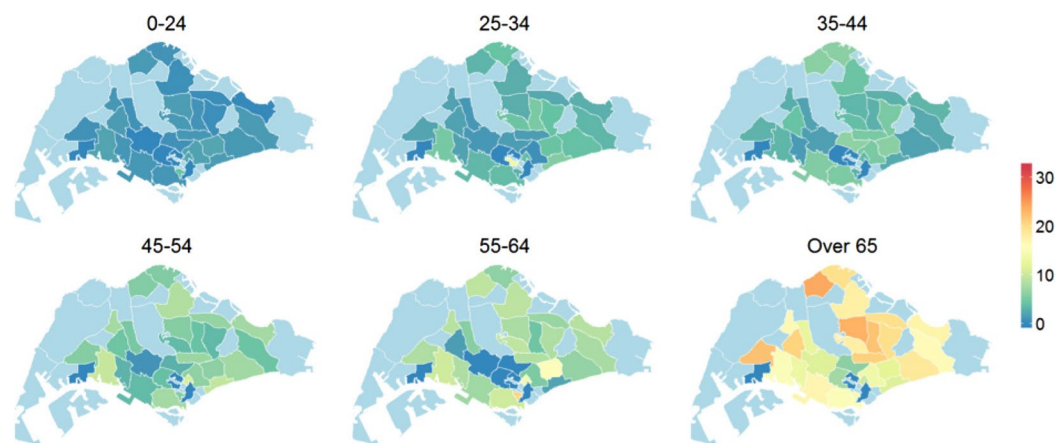


**Figure 3.**  $K$ -function plots studying clustering phenomena among various categories of reported TB cases compared with completely random spatial pattern (dashed line). (a) Annual spatial clustering curves combining all TB cases over seventeen years from 1996 to 2011, (b) Annual spatial clustering patterns of resident TB cases observed over the same time period, (c) Annual spatial clustering patterns among non-resident reported TB cases during 1996–2011, (d) Annual spatial clustering curves studying clustering of non-resident TB cases around resident TB cases reported during 1996–2011, (e) Spatial clustering curves studying clustering of resident and non-resident reported TB cases in three *high* TB (Jurong West, Bedok and Tampines) zones and *low* TB zones in 2000, (f) Spatial clustering curves showing clustering of resident and non-resident reported TB cases in three *high* TB (Jurong West, Bedok and Tampines) zones and *low* TB zones in 2010.

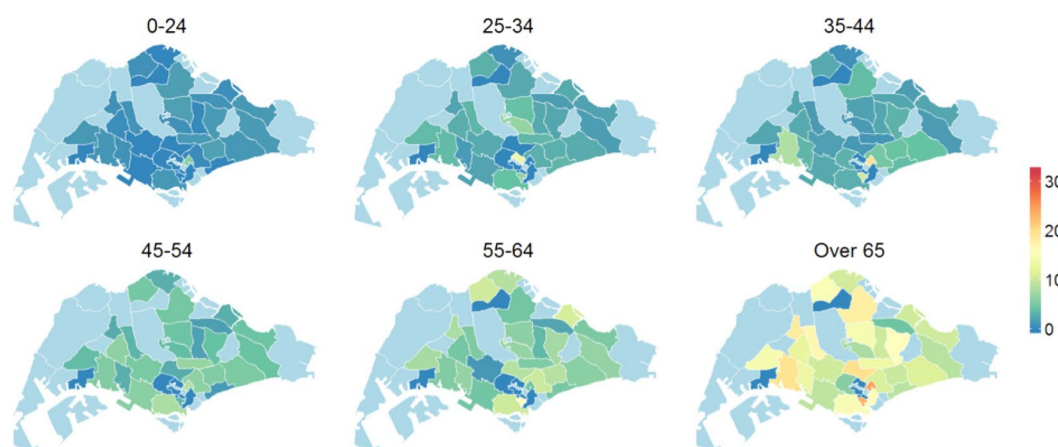


**Figure 4.** Matrix of  $p$  values from Hotelling's  $T^2$  tests. Under  $H_0$  the  $T^2$  statistic follows a central  $F$  distribution. The test is described in the Methods section. We compare the incidence rates of TB among resident Singaporeans, for planning areas of Singapore for a pairs of years. The colour *red* implies significant statistical difference in overall spatial risk between the two years while *blue* implies no difference.





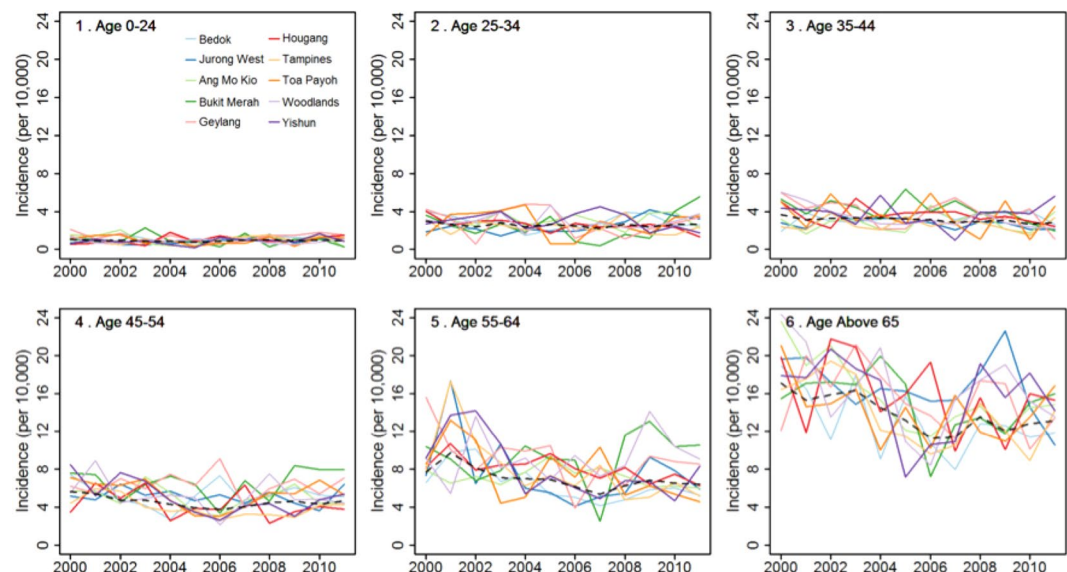
**Figure 5.** Age and zonal distribution of TB among resident Singaporeans in the year 2000. The values are adjusted for the population size and expressed per 10,000 of the susceptible population. Plots were created in statistical program R<sup>22</sup> using the package *spatstat*<sup>19</sup>. Administrative and national political coordinates of Singapore were provided, for academic research and publication, by the Singapore Land Authority (SLA)<sup>24</sup>.



**Figure 6.** Age and zonal distribution of TB among resident Singaporeans in the year 2010. The values are adjusted for the population size and expressed per 10,000 of the susceptible population. Plots were created in statistical program R<sup>22</sup> using the package *spatstat*<sup>19</sup>. Administrative and national political coordinates of Singapore were provided, for academic research and publication, by the Singapore Land Authority (SLA)<sup>24</sup>.

parts during that time. Count of reported cases and the spatial extent of distribution of cases had peaked during 1998. Since 1999 the case count and spatial range of cases had started declining, simultaneously. These coincided with the implementation of new policies on clinical and epidemiological surveillance under the Singapore TB elimination program<sup>9</sup>. During 2003–2007 the annual case count continued to remain low manifesting similar patterns of spatial variation, each year. South eastern planning areas had spatial risk of 4–6 cases per square km while rest of the city state had a low spatial intensity of 0–2 cases per km square. The period of 2008–2011 is marked by increasing annual case count together with rising spatial intensity in all administrative zones. The annual case count had started rising from its lowest levels of 1989 cases during 2002 to the highest level of 3240 in 2011. Residential development in the western planning areas also led to the unprecedented spatial risk of cases in the western planning areas.

The spatial intensity maps show that the prevalence of TB is clearly heterogeneous, with zones in the south and south east having a higher intensity of cases than most of the west of the city state. These zones are also the more established residential areas. However, the fast changing urban landscape, with increasing development of the west to accommodate a population that has increased from 3 to 5 million from 1990 to 2010, has been associated with more TB cases reported from the western planning areas. Second order spatial point pattern analyses showed strong clustering of residential address among both resident and non-resident communities. But we did not find evidence of significant residential clustering of non-resident cases in neighbourhoods with high resident incidents, indicating low probability of inter-group transmission in residential neighbourhoods. This however does not discount other forms of prospective transmission at places of occupation, business and transportation networks, an area that should be investigated in future studies.



**Figure 7.** Resident annual time series incidence for ten leading planning areas of Singapore across six different age groups from 2000 to 2010. The incidence is adjusted for underlying susceptible population.

The lack of evidence of clustering among local resident cases around foreign cases suggests that other factors are responsible for the increasing incidence of TB among residents. We found that the spatial structure of planning area level susceptible population was the primary explanatory factor underpinning spatial variation of case incidence during the time horizon of the study (1995 to 2011). It was observed that the population annual incidence rates did not show statistically significant temporal variation.

We also discovered that after stratifying by age group, at both the national and regional level, age-specific incidence among residents actually fell (in contrast to the non-age stratified incidence, which rose) over the time course of the study. One cause of the increase at the national level can therefore be attributed to the aging of the population—which, due to the low fertility and mortality rates of Singapore residents is expected to accelerate to 2050<sup>12</sup>—which has led to more individuals being of an age in which reactivation of latent TB (acquired perhaps during the 1930s to 1960s, a time when Singapore was yet to undergo its transformation from low to high income), may be more likely<sup>13</sup>.

Singapore's resident TB epidemiology is particularly relevant for other Asian countries that are developing and aging rapidly, such as China. Much like the city state, China has undergone rapid and extensive urbanization in the past three decades. China also features prominently in global terms for both TB (10% of the global cases) and aging (China being home to 23% of the global population of above 60 year olds and above<sup>14</sup>), and the growth of the latter may spur that of the former, as in Singapore.

The primary limitation of the study is that the denominator for the non-resident population is not known for Singapore, precluding incidence rates for this segment of the overall population. It is known that the size of the non-resident population has risen over the time period being studied, and now stands at 29% of the total population<sup>15</sup>, but the breakdown by age, or nationality, is not available. As a result, a full analysis of the spatial incidence rates for non-residents, matching what we did for residents, was not possible. However, the analysis of the spatial correlations between resident and non-resident cases indicated near-independence of the two groups, so the scant information on the non-resident population does not prevent a thorough assessment of the resident population.

## Conclusion

In this study, spanning 17 years of reported TB cases, we found that the major risk factor contributing to the recent rise in TB cases among Singapore residents is the burgeoning elderly population. Regions in the south and east of the city have relatively higher spatial risk due to the age distribution of susceptible population. TB incidence rates have declined among all age groups of the resident population in all areas of the city state. Reported cases display statistically significant spatial clustering within residents and non-residents, but there was no evidence of clustering between these two groups.

## Methodology

**Data.** Under the Infectious Disease Act in Singapore, all suspected and confirmed TB cases are notified to the STEP registry of the TB Control Unit (TBCU) within 72 hours of starting TB treatment and/or laboratory confirmed results. The TBCU is the national unit for treatment of TB patients, contact investigation, preventive therapy, and educational training of health care workers in Singapore. We extracted data from the TBCU on all TB cases reported to the Singapore TB Elimination Program between 1995 and 2011.

Population data were obtained from Singapore Department of Statistics<sup>15</sup> and Ministry of Health<sup>8</sup>. In particular we used planning area level population count categorized into five year age groups.

To map the spatial risk of epidemic we use spatial point process methodology. Heuristically, *events* occurring in a bounded spatial zone form a realisation of a point process if the number of location of these events, in this context, TB cases within the spatial extent of Singapore, are random processes. Realisations of a point process are called point patterns.

We define location of the  $i$ th TB case by  $(x_i, y_i)$  in the geographical window that spans the main island of Singapore, Pulau Ujong. The point pattern,  $M = \{(x_1, y_1), (x_2, y_2) \dots, (x_N, y_N)\}$ , is the set of all TB cases observed during a time period  $T$ .

We estimate first and second order moments (analogous to the mean and variance) of this point process, describing the spatial variation in risk and clustering of TB.

**Intensity function  $\lambda(\cdot)$ .** The intensity function,  $\lambda$ , at location  $(x, y)$  give the probability of observing a case in an infinitesimally small neighbourhood of  $(x, y)$ , and the map of  $\lambda(\cdot)$  over the entire spatial window represents the *spatial risk* of the event.

Mathematically  $\lambda(\cdot)$  is defined as

$$\lambda(x, y) = \lim_{\{|dx||dy| \rightarrow 0\}} \left( \frac{E[M(dx \, dy)]}{|dx||dy|} \right) \quad (1)$$

where  $E(\cdot)$  denotes the expectation. For a completely random spatial point pattern (CSR),  $\lambda(\cdot)$  measures the number of events per unit area. Consequentially,  $\int_A \lambda(x, y) dx \, dy$  gives the expected number of events in an area  $A$ . An estimate of  $\lambda(x, y)$  is given by

$$\hat{\lambda}(x, y) = e(x, y) \sum_{i=1}^n \kappa(x_i - x, y_i - y) \quad (2)$$

where  $\kappa(\cdot)$  is a bivariate Gaussian kernel with bandwidth selected by Diggle's cross-validation<sup>16</sup>, and  $e(\cdot)$  is an edge correction, specifically the translation correction<sup>17</sup> to account for the reduced effective neighbourhood size near the edges of the chosen spatial domain.

Intensity maps were created by dividing the island into ~16 000 square cells of approximately 300 m in length, and evaluating the estimated intensity function thereon.

**$K(\cdot)$  function.** The  $K(\cdot)$  function, which measures interdependence of events and acts as a surrogate for transmission of disease and/or the influence of environmental factors, is defined to be

$$K(r) = \frac{E(r)}{\lambda} \quad (3)$$

where  $E(r)$  is the mean number of events within  $r$  spatial units of another event.

Thus  $K(r)$  is a scaled measure of the additional number of events in a neighbourhood of radius  $r$  of a case, given the spatial risk. For a completely spatially random process (CSR)—with a constant intensity function—the  $K(\cdot)$  function at a distance  $r$  is  $\pi r^2$ <sup>16, 18, 19</sup>. The theoretical  $K(\cdot)$  function for an inhomogeneous spatial Poisson process with intensity function  $\lambda(s)$  is also  $\pi r^2$ .

The modified cross- $K(\cdot)$  function<sup>19, 20</sup> extends the  $K(\cdot)$  function to *two* spatial point processes observed in the same window. That is for two different point processes  $M_1$  and  $M_2$ , the cross- $K_{cM_1M_2}(\cdot)$  is a scaled average of additional number of events of  $M_2$  around an arbitrary event of  $M_1$ . In particular, we calculate the cross  $K$  function to quantify the spatial interaction between cases of TB in the resident ( $M_1$ ) and non-resident ( $M_2$ ) populations.

The non-parametric unbiased estimate of  $K(\cdot)$  function is given by<sup>18, 19</sup>

$$\hat{K}(r) = \frac{|A|}{n^2} \sum_{i=1}^n \sum_{j \neq i}^n w_{ij} I(r_{ij} \leq r) \quad (4)$$

where  $r_{ij}$  is the distance between events  $i$  and  $j$  and  $w_{ij}$  is the edge correction<sup>21</sup>. We compute the  $K(\cdot)$  function for the overall TB population and separately for resident and non-resident population. If the estimated  $K$ -function  $\hat{K}(r)$  is significantly higher than  $\pi r^2$  at distance  $r$  then the data demonstrates aggregation or clustering indicative of transmission among infected individuals or of shared environmental factors. On the other hand, a value of  $\hat{K}(r)$  significantly lower than  $\pi r^2$  indicates 'repulsion' between the two sets. All  $K(\cdot)$  curves were estimated under the assumption of inhomogeneous intensity function and compared against an inhomogeneous Poisson process. We used the function,  $Kinhom(\cdot)$ , of the spatstat package<sup>19</sup> in software R<sup>22</sup>.

**Pairwise comparison of annual spatial incidence rates.** We adapted Hotelling's two sample  $T^2$  test<sup>22</sup> to assess the hypothesis whether there were significant temporal variation in the annual spatial incidence of TB, among residents in Singapore. To compare we adjusted the case count at planning area levels to per 10,000 of the susceptible population. Let us define  $K$  to be the number of planning areas,  $T_1$  and  $T_2$  to be the two years to compare, and  $Y_k^T$  to be the *adjusted resident case count* (where the adjustment is obtained by scaling the annual incidence per 10,000 of the planning area population) in the  $k^{th}$  planning area in year  $T$ . The total adjusted resident case count in year  $T$  we denote  $N_T = \sum_{k=1}^K Y_k^T$ .

Then, assuming that individual cases arrive randomly in time (for each year) and independently in space (among any of the  $K$  planning areas), we have,

$$\mathbf{Y}^T = (Y_1^T, Y_2^T, \dots, Y_K^T)' \sim \text{Multinomial}[N_T, \mathbf{p}^T = (p_1^T, p_2^T, \dots, p_K^T)'] \quad (5)$$

where  $p_k^T$ ,  $k = 1, 2, \dots, K$ , are the population distribution of resident cases of the  $k^{\text{th}}$  planning area in  $T^{\text{th}}$  year.

$\mathbf{Y}^T$  can alternatively be expressed as the sum of  $N_T$  independent and identically distributed incidence vectors<sup>23</sup> of order  $K$ ,  $\mathbf{X}_i^T = (X_{i1}^T, X_{i2}^T, \dots, X_{iK}^T)'$ , where  $X_{ik}^T = 1$  and  $X_{i(k \neq l)}^T = 0$ , for  $i = 1, 2, \dots, N_T$ . This formally implies that the  $i^{\text{th}}$  case was reported at the  $k^{\text{th}}$  planning area and that an individual case can't be present in two planning areas (or the boundary) at the same time. Then

$\mathbf{X}_i^T \sim \text{Multinomial}[1, \mathbf{p}^T = (p_1^T, p_2^T, \dots, p_K^T)']$ ,  $i = 1, 2, \dots, N_T$ . Also,  $\Sigma = \text{Var}(\mathbf{X}_i^T)$ , is defined by

$$\text{Cov}(X_{ik}^T, X_{il}^T) = \begin{cases} p_k^T(1-p_k^T), & k = l \\ -p_k^T p_l^T, & k \neq l \end{cases}$$

Thus,  $\mathbf{Y}^T = \sum_{i=1}^{N_T} \mathbf{X}_i^T$ . Since  $\mathbf{X}_i^T$  are independently and identically distributed, for sufficiently large  $N_T$  using the central limit theorem, for a sample of random vectors, we have

$$\bar{\mathbf{X}}^T \xrightarrow{D} N_K(\mathbf{p}^T, \Sigma/N_T), \text{ as } N_T \rightarrow \infty. \quad (6)$$

**Test Statistic.** Under the above framework if  $H_0: \mathbf{p}^{T_1} = \mathbf{p}^{T_2}$  and we define

$$t^2 = \frac{N_{T_1} N_{T_2}}{N_{T_1} + N_{T_2}} (\bar{\mathbf{X}}^{T_1} - \bar{\mathbf{X}}^{T_2})' \hat{\Sigma}^{-1} (\bar{\mathbf{X}}^{T_1} - \bar{\mathbf{X}}^{T_2}), \quad (7)$$

then

$$T^2 = \frac{N_{T_1} + N_{T_2} - K - 1}{(N_{T_1} + N_{T_2} - 2)K} t^2 \sim F(K, N_{T_1} + N_{T_2} - K - 1), \text{ under } H_0. \quad (8)$$

Here,  $\bar{\mathbf{X}}^{T_j} = \frac{1}{N_{T_j}} \sum_{i=1}^{N_{T_j}} \mathbf{X}_i^{T_j}$  and  $\hat{\Sigma}_{T_j} = \frac{1}{N_{T_j} - 1} \sum_{i=1}^{N_{T_j}} (\mathbf{X}_i^{T_j} - \bar{\mathbf{X}}^{T_j})(\mathbf{X}_i^{T_j} - \bar{\mathbf{X}}^{T_j})'$  where  $j = 1, 2$ . Also,  $\hat{\Sigma} = \frac{N_{T_1} \hat{\Sigma}_{T_1} + N_{T_2} \hat{\Sigma}_{T_2}}{(N_{T_1} + N_{T_2} - 2)}$ , the pooled variance.

For the given samples we reject  $H_0$  at level of significance  $\alpha$  if

$$T^2 > F(\alpha; K, N_{T_1} + N_{T_2} - K - 1). \quad (9)$$

Alternatively we reject  $H_0$  if the p-value,

$$p = P_{H_0}[T^2 > T_{\text{Observed}}^2] < \alpha. \quad (10)$$

We computed p-values corresponding to Hotelling's  $T^2$  for all  $\binom{17}{2}$  distinct pairs (1995 to 2011) of annual TB incidence rates, among Singapore's residents. The results are shown in Fig. 3.

## References

- Daniel, T. M. The history of tuberculosis. *Respir. Med.* **100**, 1862–1870 (2006).
- Raviglione, M. C., Snider, D. E. & Kochi, A. Global epidemiology of tuberculosis. Morbidity and mortality of a worldwide epidemic. *JAMA* **273**, 220–226 (1995).
- World Health Organisation. Tuberculosis Fact Sheet. WHO (2016). Available at: <http://www.who.int/mediacentre/factsheets/fs104/en/>. (Accessed: 16th May 2016).
- Wagstaff, A. & Claeson, M. The Millennium Development Goals For Health: Rising to the Challenges (2004).
- World Health Organisation. Global Tuberculosis Report 2015. WHO (2016). Available at: [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/). (Accessed: 16th May 2016).
- WHO | Urban population growth. WHO Available at: [http://www.who.int/gho/urban\\_health/situation\\_trends/urban\\_population\\_growth\\_text/en/](http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/).
- Chee, C. B.-E. & Wang, Y. T. TB control in Singapore: where do we go from here? *Singapore Med. J.* **53**, 236–238 (2012).
- Ministry of Health, Singapore. *Communicable Disease Surveillance Report 2010* (2011).
- Chee, C. B. E. & James, L. The Singapore Tuberculosis Elimination Programme: the first five years. *Bull. World Health Organ.* **81**, 217–221 (2003).
- Bartlett, M. S. The Spectral Analysis of Two-Dimensional Point Processes. *Biometrika* **51**, 299–311 (1964).
- Wah, W. et al. Time series analysis of demographic and temporal trends of tuberculosis in Singapore. *BMC Public Health* **14**, 1121 (2014).
- Phan, T. P. et al. Forecasting the burden of type 2 diabetes in Singapore using a demographic epidemiological model of Singapore. *BMJ Open Diabetes Res. Care* **2**, e000012 (2014).
- Kline, K. A. & Bowdish, D. M. Infection in an aging population. *Curr. Opin. Microbiol.* **29**, 63–67 (2016).
- Zhang, N. J., Guo, M. & Zheng, X. China: awakening giant developing solutions to population aging. *The Gerontologist* **52**, 589–596 (2012).
- Department of Statistics Singapore. *Yearbook of Statistics 2014*. (2015). Available at: [https://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications\\_and\\_papers/reference/yearbook\\_2015/yos2015.pdf](https://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications_and_papers/reference/yearbook_2015/yos2015.pdf) (Last accessed 22nd January 2017).
- Berman, M. & Diggle, P. Estimating Weighted Integrals of the Second-Order Intensity of a Spatial Point Process. *J. R. Stat. Soc. Ser. B Methodol.* **51**, 81–92 (1989).
- Stoyan, D. & Stoyan, H. *Fractals, Random Shapes and Point Fields*. (Wiley, 1994).



18. Diggle, P. Statistical Analysis of Spatial Point Patterns. (Hodder Education, 2002).
19. Baddeley, A., Rubak, E. & Turner, R. Spatial Point Patterns: Methodology and Applications with R (2015).
20. Baddeley, A. J., Møller, J. & Waagepetersen, R. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerlandica* **54**, 329–350 (2000).
21. Ripley, B. D. The Second-Order Analysis of Stationary Point Processes. *J. Appl. Probab.* **13**, 255 (1976).
22. Hotelling, H. The Generalization of Student's Ratio. *Ann. Math. Stat* **2**, 360–378 (1931).
23. R Development Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2016).
24. Licensing of Digitised Land Information. Available at: <http://www.sla.gov.sg/Services/StreetMap/LicensingofDigitisedLandInformation.aspx>. (Last accessed: 22nd January 2017).

## Acknowledgements

S.D., A.R.C., W.W. were supported by the Centre for Infectious Diseases and Epidemiological Research, at the National University of Singapore. K.M.K.W., C.B.E.C. and Y.T.W. are supported by the Singapore TB Elimination Program. H.L.Y. was supported by the National University of Singapore and National University Health System.

## Author Contributions

S.D., A.R.C. and H.L.Y. planned and analysed data. S.D. and A.R.C. wrote the manuscript. W.W. prepared data. K.M.K.W., C.B.E.C. and Y.T.W. were in-charge of data collection and dissemination. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-00081-3](https://doi.org/10.1038/s41598-017-00081-3)

**Competing Interests:** The authors declare no competing financial interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017